Preprint
UCRL-JC-143305

Bayesian Spectroscopy and Target Tracking

C. T. Cunningham

This article was submitted to 7th International Conference on Applications of Nuclear Techniques, Crete, Greece June 17-23, 2001

May 1, 2001

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

DISCLAIMER

Bayesian Spectroscopy and Target Tracking

Christopher T. Cunningham

ABSTRACT

Statistical analysis gives a paradigm for detection and tracking of weak-signature sources that are moving among a network of detectors. The detector platforms compute and exchange information with near-neighbors in the form of Bayesian probabilities for possible sources. . This can shown to be an optimal scheme for the use of detector information and communication resources. Here, we apply that paradigm to the detection and discrimination of radiation sources using multi-channel gamma-ray spectra. We present algorithms for the reduction of detector data to probability estimates and the fusion of estimates among multiple detectors.

A primary result is the development of a goodness-of-fit metric, similar to χ^2, for template matching that is statistically valid for spectral channels with low expected counts. Discrimination of a target source from other false sources and detection of imprecisely known spectra are the main applications considered. We use simulated NaI spectral data to demonstrate the Bayesian algorithm compare it to other techniques. Results of simulations of a network of spectrometers are presented, showing its capability to distinguish intended targets from nuisance sources.

I. Introduction

For a number of years, the author has been involved with the tracking of moving radiation sources using fixed sensors. Up until now, our detectors have been simple counters. The analysis of detection events in a simple counter is straightforward, and Bayesian estimation has been used in this context for some time.[i] The Bayesian approach is well-suited to the estimation of the position of a moving object, based on imperfect information about it.[ii] Using a network of detectors to monitor the movement of sources along known routes is a particularly useful special case[iii], with several possible applications, including the tracking of shipments or the monitoring of industrial processes.

Since many different types of radiation sources may be present in such environments, we must develop a capability to discriminate the sources of interest for tracking ("true" sources) from others ("false" sources). This paper presents our initial development of such a discrimination capability. One problem is apparent from the outset, urban sources will be weak and will only be seen briefly by fixed detectors. Therefore it is imperative that the most sensitive methods possible be developed. Here, the Bayesian approach is of great value, since it can be proven to be the optimal scheme in many cases of interest. This feature is of use not only in designing discrimination algorithms, but also in

evaluating other techniques. Similarly, the ability to develop provably optimal methods for spectral classification is relevant in a larger context than tracking networks.

Another aspect of performing spectroscopy on weak sources is that there will probably be many spectral channels with few or no counts. In this situation, approximating the distribution of spectra as a multivariate Gaussian may not be valid, and this is the basis of many classification techniques. My old statistics book advises that one should group low-frequency categories together before applying a χ^2 test, obviously to the detriment of the data in these categories. Our Bayesian approach uses Poisson statistics for low-frequency events and avoids this difficulty. We shall see that is possible to define a generalization of the χ^2 metric that is similarly untroubled by low-frequency events. We will also compare the performances of conventional Gaussian techniques with our Bayesian approach.

Obviously, the proof of this classification effort is the performance of a network in which there are false sources. We have developed two general Bayesian tracking network architectures, in which the tracking function is either performed at a central node or distributed onto the detector platforms. These architectures can accommodate a variety of detector types, including spectrometers. An advantage of the approach taken here for spectrometer networks is that a great deal of data compression is performed at the detectors; only high-level probabilistic assessment of the sources present needs to be communicated among sensors, not the extensive raw data.

We simulate a very simple network in which our prototype discrimination algorithm is used for tracking a target source in the presence of a false source with a fairly similar spectrum. We find that our rejection scheme works very well: network performance against the false source alone is indistinguishable from the no-source case. Furthermore, the false-source rejection mechanism does not degrade the detection and tracking of the target source.

The paper is organized as follows:

In Section II we briefly introduce some Bayesian formalism that will be used in the remainder of the paper. A detailed discussion of decision-making is presented in the Appendices.

In Section III, we apply the decision-making formalism to spectral observations. Discriminating two known spectra, multiple known spectra, and imprecisely known spectra are the topics considered. In addition, we compare our results with those for some other commonly used techniques.

In Section IV, we present the details of data fusion in detection and tracking networks using spectrometers. We construct a simple network and perform a Monte Carlo simulation of tracking in the presence of a false source. Even though the spectra of the two sources are quite similar, we find that it is possible to reject the false source nearly completely, while maintaining nearly undiminished sensitivity for the true source.

Appendix A presents a statistical interpretation of Bayesian decision-making.

Appendix B proves the optimality of the Bayesian likelihood ratio as a decision variable for detection.

II. Bayesian decision-making formalism

Suppose one measures a system to determine the validity of a number of hypotheses which might describe it. For example, the system could be a radioisotope sample, the measurements could be the multi-channel readings of a spectrometer, and the hypotheses could be the various possibilities for composition of the sample. Or, the system could be an environment in which a particular radiation source or sources might be moving, the measurements could be the outputs of several radiation detectors and the hypotheses could be the type and current location of a source, plus the null hypotheses that no source is present.

A full treatment of the Bayesian decision-making scheme is presented in the Appendices. In summary, the fundamental quantity for this analysis is the **posterior probability $P_M(h)$** of hypothesis h for measurements M, $P_M(h)$, which is the fraction of elements for which h is true in a statistical ensemble of elements for which M is true. $P_M(h)$ evolves for the i^{th} measurement of value m_i according to Bayes' Theorem

$$P_i(h) = \frac{p(m_i \mid h)\, P_{i-1}(h)}{\Sigma_{h'}\, p(m_i \mid h')\, P_{i-1}(h')} \qquad (1)$$

where $p(m_i \mid h)$ is the conditional probability, or **likelihood**, of obtaining the measurement value m_i if h is true.

In a problem with an obvious null- or reference hypothesis h_0, such as the "no-target" or "background-only" case for radiation detection, one defines the **probability quotient $Q_M(h) \propto P_M(h) / P_M(h_0)$** and the **likelihood ratio $r(m_i \mid h) \equiv p(m_i \mid h) / p(m_i \mid h_0)$**. Then, from Bayes' Theorem

$$Q_i(h) = r(m_i \mid h)\, Q_{i-1}(h) \qquad (2)$$

If the measurement likelihoods are independent, we may define an **overall likelihood ratio $R(h) \equiv \Pi_i\, r(m_i \mid h)$** that determines the posterior probability quotient Q(h) from its initial value $Q_0(h)$:

$$Q_i(h) = R(h)\, Q_0(h) \qquad (3)$$

The overall likelihood ratio R(h) is proven in the Appendix to be an optimal decision variable for the hypothesis h, leading to maximum **probability of detection** (h is declared true for measurements for which h is true), with minimum **probability of false**

alarm, (h declared true for measurements for which h_0 is true). Any monotonic function of R(h) will also be an optimum decision variable, such as the **logarithmic likelihood ratio** $\Lambda(h) \equiv \ln R(h)$.

III. Single-detector spectroscopy

III.A. Likelihoods for known spectra

A spectrometer measures counts c_i in a number of energy channels i. It might be measuring background (the null hypothesis **b**), for which the expected counts in each channel is b_i; or else it might be measuring a known source in addition to background (the source-present hypothesis **s**), for which the expected counts are $s_i > b_i$. Here, we use **b** (or **s**) to denote, interchangeably, the set of expected values for background and the hypothesis that those values are correct. The likelihood for a measurement of c counts if an average s is expected is

$$p(c \mid s) = e^{-s} s^c / c! \tag{4}$$

This may be written using a measure, $\chi^2_P(c,s)$ of the "distance" from the observed counts c to the expected counts s, assuming a Poisson distribution. We use this notation to emphasize its similarity to the conventional χ^2-measure for Gaussian distributions. (See next section)

$$p(c \mid s) = p(c \mid c)\, e^{-\frac{1}{2}\chi^2_P(c,s)} \tag{5a}$$

$$p(c \mid c) = e^{-c} c^c / c! \tag{5b}$$

$$\chi^2_P(c,s) = 2c \ln(s/c) + 2(s-c) \tag{5c}$$

So, the overall likelihood is

$$p(\mathbf{c} \mid \mathbf{s}) = [\Pi_i\, p(c_i \mid c_i)]\; e^{-\frac{1}{2}\chi^2_P(\mathbf{c},\mathbf{s})} \tag{6a}$$

$$\chi^2_P(\mathbf{c},\mathbf{s}) = \Sigma_i\, 2c_i \ln(s_i/c_i) + 2(s_i-c_i) \tag{6a}$$

The logarithmic likelihood ratio $\Lambda(s)$ is

$$\Lambda(\mathbf{s}) = -\tfrac{1}{2}\left[\chi^2_P(\mathbf{c},\mathbf{s}) - \chi^2_P(\mathbf{c},\mathbf{b})\right] = \Sigma_i\, c_i \ln(s_i/b_i) - (s_i-b_i) \tag{7}$$

This ratio may be determined "count-by-count", rather than using accumulated counts in energy channels: In the limit of a large number of spectral channels, the logarithm of the likelihood ratio has a particularly simple form

$$\Lambda(\mathbf{s}) = \Sigma_c \ln(s/b)_c - (S-B) \tag{8}$$

where S and B are the total source-present and background-only counts expected during the interval of observation and $(s/b)_c$ is the ratio of source to background count rates expected for count c (on the basis of its energy, time of arrival, etc.). This formula has the simple interpretation that all spectrometer counts are not equivalent for distinguishing the source from background; rather they are weighted by an increasing function of the source-present / background-only ratio of expected counts.

III.B. Comparison to other multivariate measures

It is common to approximate the distribution of possible measurement values as a multivariate Gaussian. For example, Mitchell[iv] uses the observed number of counts in a channel as an empirical measure of its variance for a χ^2 test of goodness-of-fit:

$$\chi^2_E(c,s) = (s-c)^2/c \quad : \quad c > 0 \tag{9a}$$

We term this the **"Empirical" measure**. This expression is not defined for channels with $c_i = 0$. Mitchell attempts to correct the variance for low-count channels by averaging among nearby channels, a process that we will not attempt to reproduce here. Rather, to illustrate the importance of such channels we consider two alternatives:

"Empirical(-)": $\chi^2_{E-}(c,s) = 0 \ : c = 0;$

"Empirical(+)": $\chi^2_{E+}(c,s) = 2s \ : c = 0.$

The latter alternative follows from the Poisson distribution (4).

A corresponding discriminant for source vs. background is, by analogy to $\Lambda(\mathbf{s})$:

$$\begin{aligned} \Lambda_E(\mathbf{s}) &= -\tfrac{1}{2} [\chi^2_M(\mathbf{c},\mathbf{s}) - \chi^2_M(\mathbf{c},\mathbf{b})] \\ &= \Sigma_i [c_i - \tfrac{1}{2}(s_i + b_i)] (s_i - b_i) / c_i \quad : \quad c_i > 0 \end{aligned} \tag{9b}$$

This approach invites comparison with a multivariate **Gaussian measure.** The likelihood of measuring c counts for a Gaussian distribution with $\mu = \sigma^2 = s$ is

$$p_G(c \mid s) = \frac{e^{\frac{-(c-s)^2}{2s}}}{\sqrt{2\pi s}} \tag{10a}$$

Thus, a Gaussian approximation for the overall likelihood for a sequence **c** of multi-channel measurements with expected values **s** is

$$p_G(\mathbf{c} \mid \mathbf{s}) = \Pi_i (2\pi c_i)^{-\frac{1}{2}} e^{-\frac{1}{2}\chi_G^2(\mathbf{c},\mathbf{s})} \tag{10b}$$

$$\chi_G^2(\mathbf{c},\mathbf{s}) \equiv \Sigma_i (c_i - s_i)^2 / s_i + \ln(s_i / c_i) \tag{10c}$$

$$\Lambda_G(\mathbf{s}) = -\tfrac{1}{2}\,[\chi_G^2(\mathbf{c},\mathbf{s}) - \chi_G^2(\mathbf{c},\mathbf{b})]$$
$$= \tfrac{1}{2}\Sigma_i\,[(c_i^2 - s_i b_i)\,(s_i - b_i) / s_i b_i \; - \ln(s_i/b_i)] \qquad (10d)$$

These formulae use a modification to the conventional χ^2 goodness-of-fit metric to make it better reflect the measurement likelihood. The Gaussian approximation to a Poisson distribution is not very good for low numbers of observed counts and expected values, but since it is well-defined, we will not attempt to correct it (as we did χ^2_E).

Gosnell[v] and Keillor[vi] advocate use of discriminants tied to the principal axes of the covariance matrix σ^2_{ij} of all possible spectral measurements. If there are N spectral types k each with weight w_k and expected spectrum s_{ki} then the covariance matrix is

$$\sigma^2_{ij} = \Sigma_k\, w_k\, s_{ki}\,(s_{kj} + \delta_{ij}) - \mu_i\,\mu_j \qquad (11a)$$

$$\mu_i = \Sigma_k\, w_k\, s_{ki} \qquad (11b)$$

Gosnell employs the **Mahalanobis measure**[vii] D^2, defined on a reduced set of principal axes a,

$$D^2(\mathbf{c},\mathbf{s}) = \Sigma_a\,(c_a - s_a)^2 / \sigma^2_{aa} \qquad (12a)$$

We found it necessary to use many axes to get good discrimination performance. If it is taken over all axes:

$$D^2(\mathbf{c},\mathbf{s}) \rightarrow \Sigma_{ij}\,(c_i - s_i)\,(\sigma^2)^{-1}{}_{ij}\,(c_j - s_j) \qquad (12b)$$

The corresponding Mahalanobis measure for source vs. background is

$$\Lambda_M(\mathbf{s}) = -\tfrac{1}{2}\,[D^2(\mathbf{c},\mathbf{s}) - D^2(\mathbf{c},\mathbf{b})] = \Sigma_{ij}\,[c_i(s_j - b_j) - \tfrac{1}{2}(s_i s_j - b_i b_j)]\,(\sigma^2)^{-1}{}_{ij} \qquad (12c)$$

Keillor uses the simple **"count error"** measure

$$\varepsilon(\mathbf{c},\mathbf{s}) = [\Sigma_a\,(c_a - s_a)^2]^{\frac{1}{2}} \qquad (13a)$$

on a reduced set of principal axes a. Taken over all principal axes, this becomes simply the sum of square errors in observed counts:

$$\varepsilon^2(\mathbf{c},\mathbf{s}) \rightarrow \Sigma_i\,(c_i - s_i)^2 \qquad (13b)$$

So a corresponding discriminator would be:

$$\Delta E^2(\mathbf{s}) = -\tfrac{1}{2}\,[\varepsilon^2(\mathbf{c},\mathbf{s}) - \varepsilon^2(\mathbf{c},\mathbf{b})] = \Sigma_i\,[c_i - \tfrac{1}{2}(s_i + b_i)]\,(s_i - b_i) \qquad (13c)$$

III.C. Differentiating two known spectra

The problem of deciding between two expected spectra on the basis of observed counts is equivalent to that of detecting a known source above a known background. The Bayesian prescription is to use the likelihood ratio, or equivalently $\Lambda \equiv \ln R$, as the decision variable. This gives maximum probability of detection for constant probability of false alarm, and vice versa. (Or, equivalently minimum error in classification of the first spectrum for constant error for the second, and vice versa.)

As a numerical experiment, we use a combination of a simulated NaI spectrum for an exemplary shielded isotope with multiple lines (source "S") and a typical NaI background spectrum of fixed strength (background "B"), as provided by Keillor[viii]. These spectra are shown in Figure 1. (The broadness of the lines is largely due to the shielding assumed.)

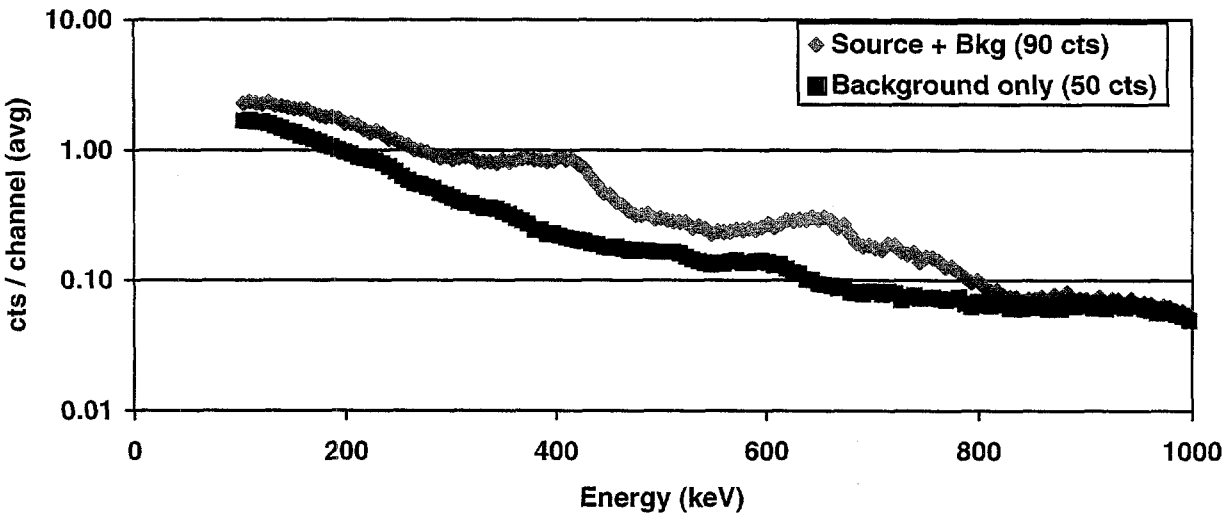

Fig. 1. Trial spectra for discrimination. The background is a typical NaI detector spectrum of 50 counts total. The source spectrum is a Monte Carlo simulation of a particular shielded isotope. The source strength shown is 40 counts above background.

The discriminants $\Lambda_M(\mathbf{s})$, $\Lambda_G(\mathbf{s})$, $\Lambda_D(\mathbf{s})$ may be considered to be approximations for the logarithmic likelihood ratio, and may therefore be compared to it as decision variables. The measure $\Delta E^2(\mathbf{s})$ is also a reasonable decision variable; it is very like $\Lambda_M(\mathbf{s})$ except for its normalization. Additionally, since the total expected counts is greater for the source-present spectrum than for background, we can use the total observed counts as another decision variable.

To assess the performance of the various discrimination methods, we drew a large number (10,000) of trial spectra from the averages given above and calculated the several

discriminants for each. The decision threshold for each method was adjusted to give equal misclassification errors (equal probabilities of detection leakage $P_{DL} \equiv 1-P_D$ and false alarm P_{FA}), and the inverse of the resulting error was taken as the measure of performance for each method. Results are given in Figure 2.

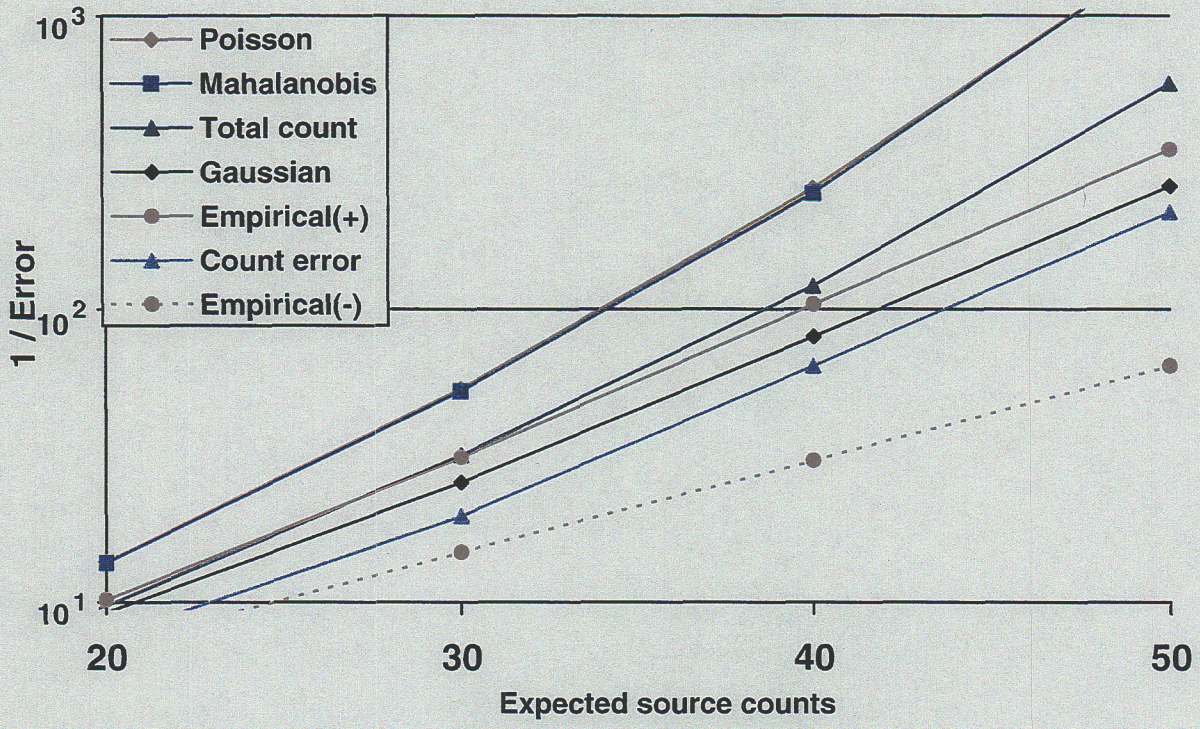

Fig. 2. Results for the various discrimination methods. The figure of merit is the inverse of the misclassification errors (probabilities of detection leakage and false alarms) observed for a large number of Monte Carlo spectral instantiations, with decision thresholds set for equal errors ($P_{DL}=P_{FA}$). Performance improves for all methods as the source strength increases. (Background is held constant.)

The big surprise is the excellent performance of the Mahalanobis measure, which is indistinguishable from that of the Poisson measure. We verified this performance of the Mahalanobis scheme in several other 2-spectra problems, always obtaining very nearly optimal results. Nonetheless, the Mahalanobis measure relies on a multivariate Gaussian approximation for the distribution of measurement results, and that approximation is not valid for small numbers of counts. Apparently, the principal-axis approach takes advantage of the similar character of many of the spectral channels to construct axes along which the expected numbers of counts are not small; so that the Gaussian approximation becomes mroe nearly valid. The main drawback with the technique is that its computational complexity is $o(N^2)$ for N channels, compared to o(N) for the optimal Poisson scheme.

These optimal schemes outperform the total counts as a discriminator, since they focus attention on regions of the spectrum with a large signal-to-noise ratio, rather than treating all counts as equivalent for discrimination, a well-known phenomenon. The effect is not strong, however, since the source spectrum exceeds background over a large number of channels, rather than in just a few sharp lines. (The detection of sources with sharper should benefit more from these techniques; as we will see in Section IV.C.)

The Gaussian and Empirical schemes take the discrimination problem channel-by-channel, and so suffer from the inadequacy of the Gaussian approximation for small numbers of counts. The enormous difference made by the zero-count correction to the Empirical scheme illustrates that low counts are at the root of the problem. These methods do not even perform as well as the simple total-count measure.

Another poor performer is the "Count error" metric. It is dominated by channels in which the variance in counts is large (the low-energy channels), regardless of whether they provide much information for discrimination (as they do not for our example source).

III.C. Discriminating three or more known spectra

In a radiation detection problem, it may not be adequate to determine that a spectrum differs from background: There might be other sources in the vicinity besides the target one is attempting to detect. Suppose we are attempting to detect the shielded source considered above in an environment in which a false source with a similar spectrum is present. For an example, we select a false source from Keillor' simulations that has lines at roughly the same energies as the example source and that matches S in total counts. See Figure 3.

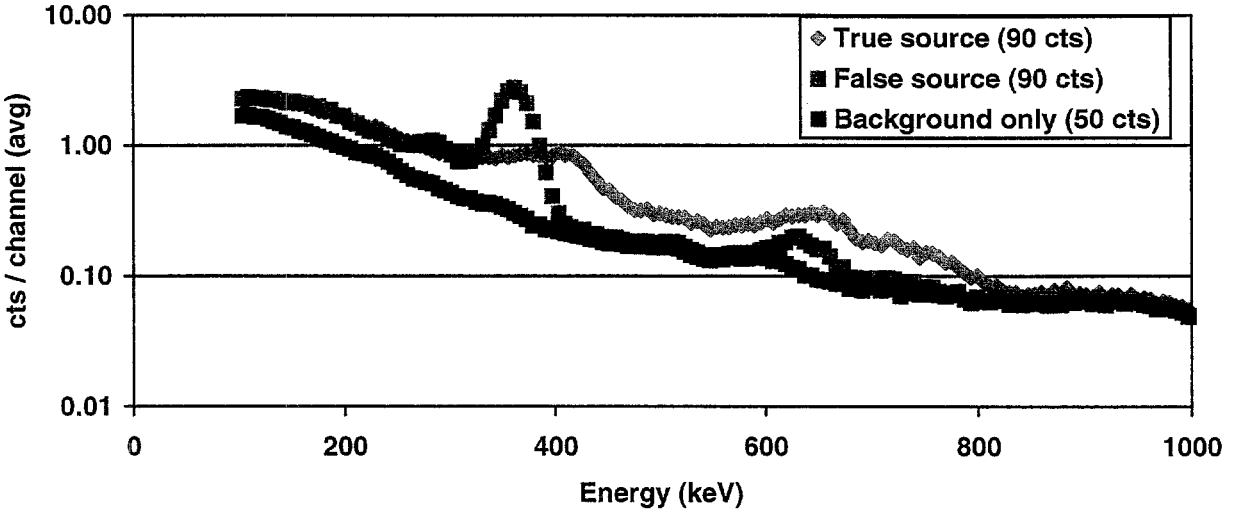

Fig. 3. Monte-Carlo simulation for the spectrum of a false source with lines of similar energies to the source of Fig. 2

Suppose that the strengths and spectra of both sources are known and that it is also known that the false source, rather than pure background, is appears a fraction P_{FS} of the time.

By equation (A8) of Appendix A, the 'corrected' likelihood ratio R` for detecting a target source S against a "combined background" in which a false source FS is present is

$$R`(S) = \frac{R(S)}{1-P_{FS} + P_{FS}\, R(FS)} \qquad (14)$$

where R(S) and R(FS) are the likelihood ratios against the "pure background", as above. This is a truly "Bayesian" measure because it uses more information than the simple measurements; namely, the probability for the false source P_{FS}.

More conventionally, we might compare the likelihood for the target source with the most likely alternative for the combined background, either pure background or the false source. In this case the corrected likelihood ratio is

$$R"(S) = \frac{p(\mathbf{c}|S)}{\max[\, p(\mathbf{c}|Bkg),\, p(\mathbf{c}|FS)\,]} = \frac{R(S)}{\max[\, 1,\, R(FS)\,]} \qquad (15)$$

We term this the "Poisson measure" to differentiate it from the Bayesian measure above, it uses Poisson statistics but does not make use of knowledge of P_{FS}.

This technique can be used with the other measures as well, using their approximations for the likelihood ratio.

We performed a Monte Carlo simulation similar to that for Section III.B but in which a false-source spectrum was used with $P_{FS} = 0.05$. The resulting spectra were interpreted using all the discrimination techniques discussed above. Results are shown in Figure 4.

The figure shows that the presence of the false source can have a major detrimental impact on the ability to detect the target source if its presence is not anticipated. In this case, nearly every appearance of the false source results in a false alarm, since its spectrum is unlike the background and quite similar to that of the source; therefore, the probability of false alarm does not fall much below P_{FS}. However, if the false source is anticipated, the discrimination can be nearly as effective as if there were no false source, using either the Bayesian technique or the Poisson or Mahalanobis maximum likelihood techniques. The performances of the Gaussian and Empirical techniques are very sub-optimal with a false source. (Compare Fig. 2.)

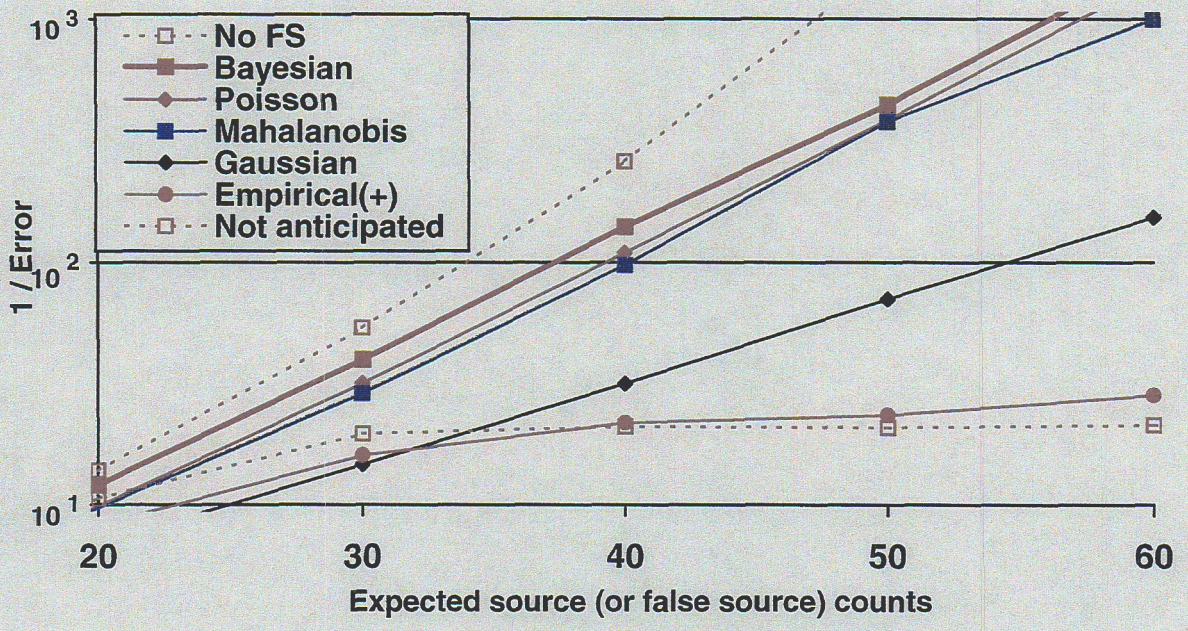

Fig. 4. Comparison of discrimination techniques with an infrequent false source. Parameters are similar to Fig. 2. For "No FS" the false source is not present, not anticipated, and the Poisson scheme is used. (This result is also given in Fig. 2.). For the remaining curves, the false source is present with $P_{FS} = 0.05$. The "Bayesian" curve assumes knowledge of P_{FS}. The "Not anticipated" curve applies Poisson discrimination, but with no provision for the false source. The remaining curves decide between pure background and a false source on the basis of maximum likelihood, using various approximations for the likelihood.

The difference between the Bayesian and Poisson maximum-likelihood techniques is statistically significant, and this lends empirical support to the optimality of the Bayesian technique. However, in this case there is not much practical difference between the effectiveness of the two.

There is a discernable difference between the Poisson and Mahalanobis techniques. We speculate that including more than two sources in the covariance matrix makes it inappropriate for any pair-wise discrimination test. That probably didn't matter very much in this case because the target-source and false-source spectra are very similar.

III.E. Imprecisely known spectra

The above examples assumed perfect knowledge of the source and false-source spectra, which would not be the case in practice. Our uncertain knowledge of a spectrum of type S can be quantified as an *a priori* statistical distribution $P`_S(\mathbf{s})$, where **s** is a set of expected channel counts. Then likelihood ratio of S for a given set of measurements is, by (A8),

$$R(S) = \int R(\mathbf{s})\, P`_S(\mathbf{s})\, d\mathbf{s} \qquad (16)$$

For example, suppose that the shapes of the source and false-source spectra are known exactly, but not their magnitudes, and that these (magnitudes) are distributed in some known fashion. We can then repeat the above numerical experiment, randomly choosing strengths from the assumed distributions and using (16) to determine the likelihoods.

However, it would seem that there should be little doubt about the magnitude of a spectrum once it has measured. The source or false-source strength can be estimated to be the difference between the observed counts and the known expected background counts. This estimate could then be used in determining the likelihood ratios for discrimination. A similar approach would be to vary the expected source strength to find the maximum likelihood ratios for a given observed spectrum, then use those maximum values

We repeat the analysis of Figure 4, for discrimination with a false source that is present 5% of the time. In this case, our imprecise knowledge of source strength is simulated by varying the source and false source strengths are randomly around an average value. (This average is the same for both the source and the false source, as for Figure 4.) The actual strengths are uniformly distributed between 0.5 and 1.5 of these averages (30% std. error). As in Figures 2 and 4, the decision threshold is chosen for equal discrimination errors, and the inverse of this error is used as the measure of performance. Figure 5 shows the performance for various discrimination techniques as a function of the averages of the true source or false source strengths. (They are equivalent for this test.)

The variation in source strength appears to be a major determinant of system performance. Much of this effect is not due to imprecise knowledge, however, but simply to the presence of relatively weak sources for which the discrimination errors are large. For example, sources with strengths between 25 and 75 counts will be present in this model for an average source count of 50. The data presented in Figure 4 allows us to estimate an overall discrimination error of 0.007 with perfect knowledge of source strengths uniformly distributed in this range, and this is not very different than the error of 0.011 achieved by the Bayesian method here.

More significantly, the figure shows that the Bayesian method is significantly better than the others for dealing with uncertainties in source strength. Surprisingly, the next best method is that for which the average count values are used to determine the likelihoods, ignoring the variation in expected counts entirely. The two schemes that adjusted the

expected counts based on observations performed worse, and adjusting counts for maximum likelihood fared the worst.

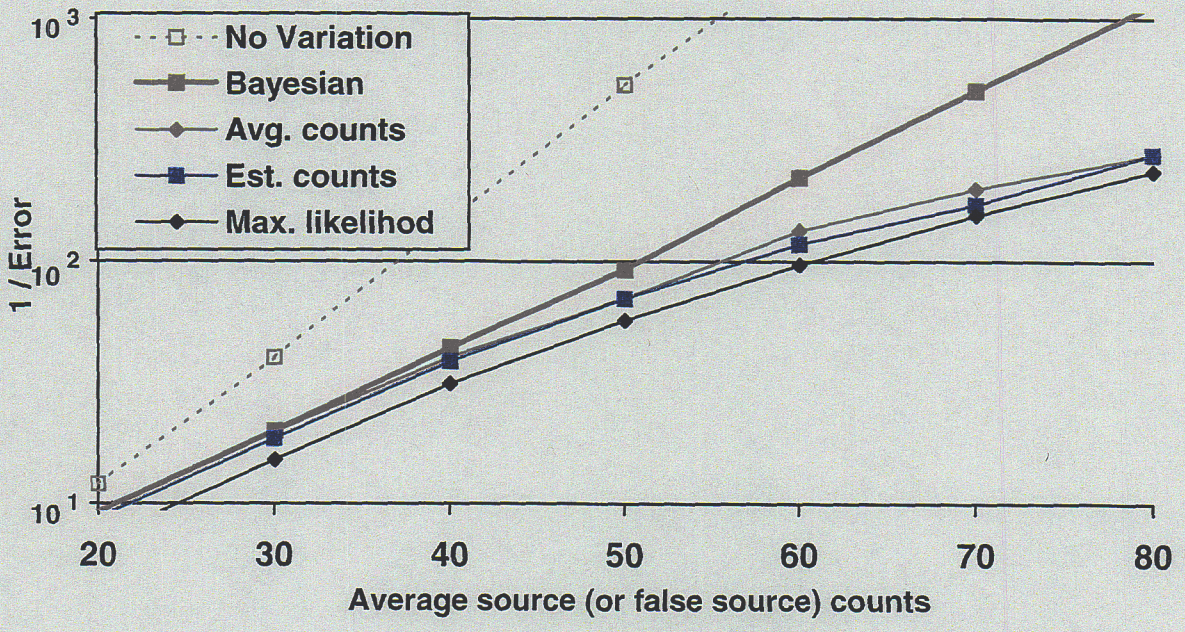

Fig. 5. Discrimination performance for imperfectly known source and false source strengths. These vary from 0.5 to 1.5 of their average values, except for the "No Variation" case, for which the expected strengths are perfectly known. The "Bayesian" curve utilizes likelihood ratios that are linear combinations over possible strengths. The "Average counts" curve uses the known averages to calculate the likelihoods, ignoring the variation in strengths. The "Estimated counts" curve uses the difference between the observed and the expected background counts as an estimated source strength. Finally, the "Maximum likelihood" curve uses the maximum values of the likelihood ratios found by varying the expected strengths for a given observed spectrum.

It may seem heuristically appealing, in this context and others, to try to determine the maximum likelihood possible consistent with observations by adjusting system parameters, such as the source strengths in this case, and then use that maximized likelihood to validate assumptions about the system. That may well be an incorrect approach, as seen here. Note that we were able to use a similar approach for the composition of the background in the previous section without much impact on performance, but these results cast doubt on that technique as well.

IV. Spectrometer networks

IV.A. Bayesian Tracking Networks

Our primary focus for this work is to detect, identify, and track radiation sources as they move within a network of spectrometers. However, a rather general prescription can be given for a detector network to perform these functions, independent of details of the specific detectors. We describe two such network paradigms, one in which the tracking calculation are idone centrally, and the other in which it is done locally at the detector platforms. These paradigms are applicable to any type of sources moving through a network of detectors, provided that the distance between detectors is greater that a detector's characteristic range.

In such a tracking network, we determine Bayesian probability quotients Q(x, t, s) as functions of location, time, and source type. These quantities evolve due to our knowledge of the source's movement and also due to our interpretations of detector reports.

The source is idealized as moving between nodes in a "road network". Such a network might represent other terrain as well, such as paths of objects in a factory, or the flow of commerce. Each node x, therefore, has only a few well-defined adjacent nodes y in the network. We use a movement model in which the movement probabilities p(x, y) for a source to move from x to y in unit time, with $p(x, x) = 1 - \Sigma_{y \neq x} p(x, y)$. These movement probabilities and the measurement likelihoods represent the bulk of our *a priori* knowledge about the system. Such a first-order movement scheme might seem overly simplistic, but it has performed very well in fairly demanding problems, such as sources moving within a city.

If P(x, t, s) is the Bayesian probability for source s at x, t, and there are no intervening measurements, then

$$P(x, t+1, s) = \Sigma_y \, p(x, y) \, P(y, t, s) \tag{17a}$$

Since we assume that there are no measurements during this movement, P(t+1,n) = P(t,n), for the no-source hypothesis n. Thus, since $Q(x, t, s) \equiv P(x, t, s) / P(x, t, n)$,

$$Q(x, t+1, s) = \Sigma_y \, p(x, y) \, Q(y, t, s) \tag{17b}$$

We have already determined the effect of measurements on the probability quotient, so, for the measurements **M**(t) made at time t

$$Q(x, t_+, s) = R(\mathbf{M}(t), x, s) \, Q(x, t_+, s) \tag{18}$$

where R(**M**(t), x, s) is the likelihood ratio for a source s at x, due to all measurements made at time t.

An important feature for radiation detection in a tracking system, in which distances are large compared to the ranges of the detectors, is that most measurement likelihoods are indistinguishable from those for background. Specifically, we can assume that for the likelihood of observing measurement m for s at x, p(m | x, s), is equivalent to the likelihood of observing m with no source in the network, p(m | n), unless x is the detector at which the measurement m is made. Therefore, if x is not a detector location,

$$Q(x, t_+, s) \approx Q(x, t_-, s) \tag{19a}$$

and if x is a detector location d

$$Q(d, t_+, s) = R(\mathbf{M}_d(t), d, s)\, Q(d, t_-, s) \tag{19b}$$

where $\mathbf{M}_d(t)$ includes only those measurements made at d at time t. It is precisely this reduction in complexity that motivates use of the likelihood ratio and probability quotient. This completes the specification of the evolution of Q(x, t, s).

In Appendix A we noted that the probability quotient Q is an optimal decision variable for detection. Thus, Q(x, t, s) is an optimal decision variable for detecting source s at position x at time t, and $Q(t, s) = \Sigma_x\, Q(x, t, s)$ is an optimal decision variable for detecting source s anywhere in the network.

The probability that source s is at (x, t) assuming that s is in the network ("| in") is proportional to its Bayesian probability of being at (x, t) and therefore may be written

$$p(x, t, s \mid in) = Q(x, t, s) / Q(t, s) \tag{20}$$

This quantity and Q(t, s) give important location and detection information about possible sources and develop with time in response to measurements.

IV.B. Distributed Tracking Networks

The tracking scheme is very easy to implement with distributed processing, in which the primary probability calculations are done locally at the detector platforms. Such distributed processing may be advantageous because information flow to a central site is reduced and because information about a particular source travels along with it in the network. Thus, a shipment of nuclear material, for example, could be identified at the moment it passed the last detector in a sequence of observations, utilizing all available information and not requiring notification to and response from a central decision node.

The only change required to the centralized scheme described above is the movement model. In the distributed scheme we do not consider the intermediate locations x, but only the detector locations d. Thus, rather than the time-stepping movement model (17a), we have an array of movement probabilities between detectors: p(d, d`, δt) is the

probability that if a source is at detector d at a given time it will be at adjacent detector d` after a delay $\delta t$. Finding p(d, d`, δt.) from p(x, y) is straightforward.

In the distributed scheme, when detector d` computes Q(d`, t`, s) it sends that quantity to an adjacent detector d, provided that Q exceeds some threshold for significance, to reduce communications traffic. The calculation at d of new values of Q proceeds as

$$Q(d, t, s) = R(\mathbf{M}_d(t), d, s)\,[\Sigma_{d`,\delta t}\, p(d`, d, \delta t.)\, Q(d`, t-\delta t, s)] \qquad (21)$$

Detector d then sends Q(d, t, s) to its neighbors and the process continues. Thus, the detector platforms, always "know" the latest values of Q(d, t, s), which is the basis for decision-making by them. (At least for the first detection decisions, after that, more global considerations might become important.)

Significant values of Q(d, t, s) are also sent to the central site (or other global monitoring sites). Here, they are propagated through the network using the time-stepping movement model (17a). The probability map p(x, t, s | y) and overall probability quotient Q(t, s) are, thus known at the central site. The central site might advise the detectors of the values of Q(t, s), to revise their notions of a "significant" value of Q(d, t, s).

IV.C. A Spectrometry Example

Suppose that the target source and false source considered in Section III move in the simple network of detectors shown in Figure 6.

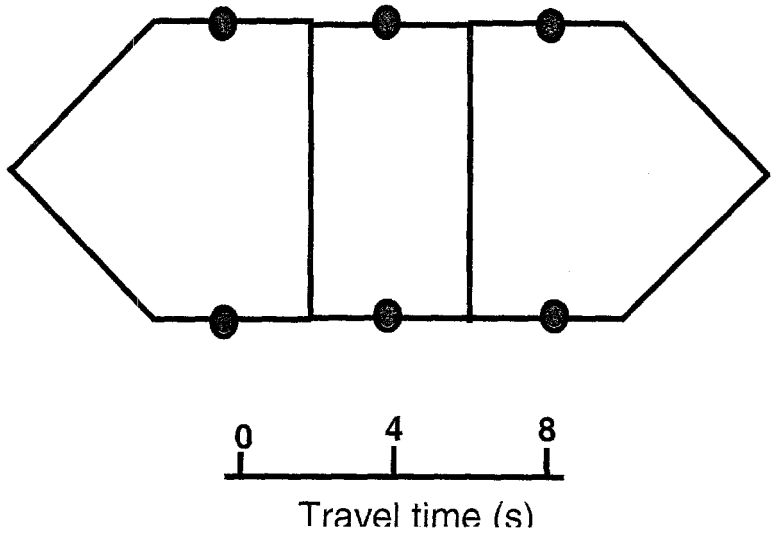

Fig. 6. A simple network. Sources move in either direction along the paths.

The motion of the sources was modeled as stochastic, with an 80% chance of moving each second.

We assumed that when a source passed a sensor it would make a measurement with a total background of 50 counts. We considered average source strengths (above background) of 10, 15, 20 counts. We assumed that these average strengths were known accurately, but the actual strengths at the time of measurement were taken to vary randomly between 50% and 150% of the average, as in Section III.E. The discrimination technique of Section III.C, eq. (14) was used to suppress observations of the false source.

Results for a typical Monte Carlo simulation for movement within this network are given in Figure 7.

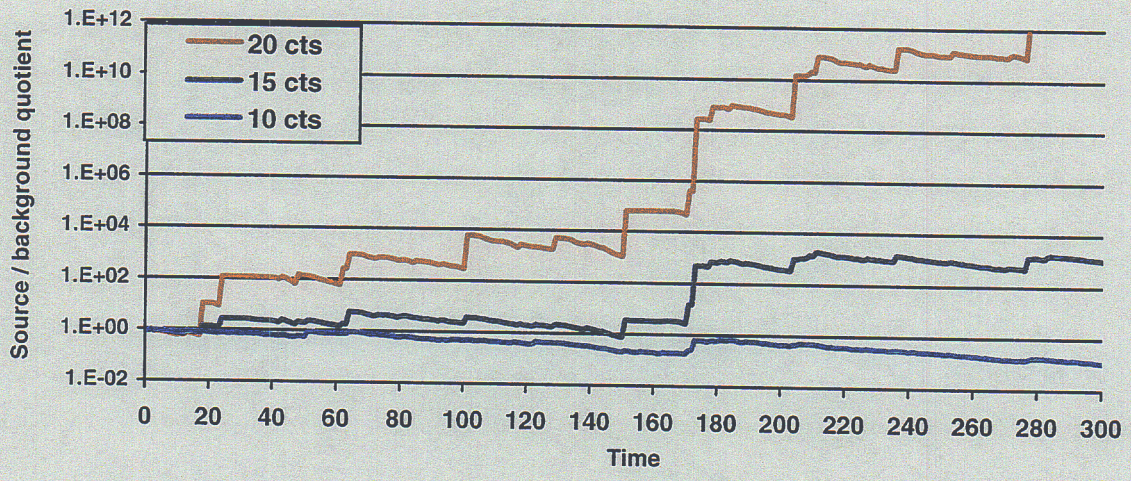

Fig. 7. Results for a typical Monte Carlo simulation of network performance as source and false source move in the network. The curves show results for different average source strengths. (The source and false source had the same average strength, as in Section III.) To facilitate comparison, the same random number sequence was used throughout.

Three behaviors are evident. For the larger source strength (20 counts), the source / background quotient Q grows steadily. The network data fusion accumulates the likelihood ratios, and they are sufficiently greater than unity that confidence that a source is present (Q) increases. The smaller strength (10 counts) is not adequate to offset the uncertainty in the target's movement. In this case, the source / background quotient remains or returns near to its initial value of unity. (The measurements are so ambiguous that they do not increase our confidence.) The middling strength (15 counts) is near an interesting transition: Q does not generally increase, but nor does it decay toward its initial value after a fortuitous detection event. In this case the likelihood ratios are adequate, barely, to offset the uncertainty in motion.

This notion can be quantified to estimate network performance. The factor that causes the probability quotient to degrade between sources is the matrix of transition probabilities p(d, d`, δt); a measure of this degradation is its average

$$< p(d, d`, \delta t)> = N_d \, \Sigma_{d,d`\delta t} \, p(d, d`, \delta t)^2$$

For our network, this average is about 0.1.

The opposing factor is the mean of the likelihood ratio when the source is present. We saw in Fig. 2 that total counts is not too bad an approximate descriptor of detection performance for the source. For a single spectral channel of total source-present counts S and background counts B an appropriate average of the likelihood ratio is $(S/B)^S e^{B-S}$, by eq. (4). For B=50 and S=65, this average is about 8, and the product of the two factors is about unity, which shows that this is near the transition source strength, as we observed above.

As a measure of tracking performance, we take the conditional location probability measured at the source, p(x, t, s | det) = Q(x, t, s) / Q(t, s). Since the typical distance between sensors is about 10 units, a probability of ~0.1 may well indicate that the source is being successfully tracked. (If the source's probability is spread uniformly in the network p ~ 0.01). This probability is shown as a function of time in Figure 8 for the three trial source strengths.

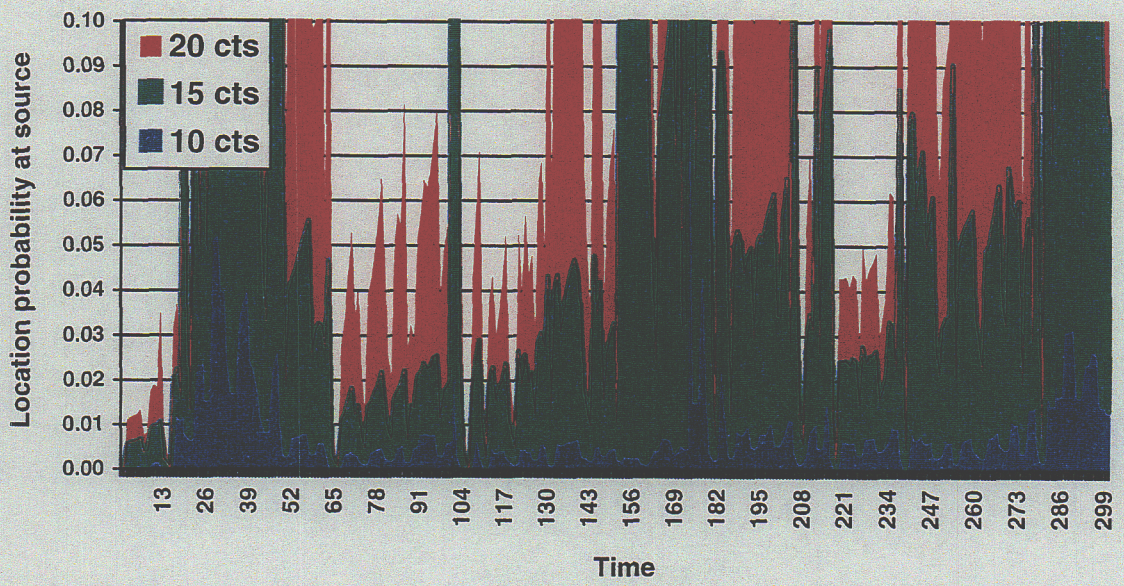

Fig. 8. Location probability at the source for the sample network run and three source strengths.

The figure shows that for a source strength of 20 counts, the source is tracked a significant fraction of the time, while for a strength of 10 counts it is hardly ever tracked. The strength of 15 counts represents a transition case in this regard as well.

These results included a false source in the network and false-source rejection in the network algorithm, which performed very well. To illustrate this performance, we replay this run with a number of different source conditions, with and without false-source rejection. Results are given in Figure 9.

The figure shows that with false-source rejection, the behavior of Q during the run is indistinguishable from background if only the false source is present. Similarly, the behavior is as for the true source alone if both the true source and the false source are present. Significantly, the false-source rejection procedure does not weaken the detection capability. Without false source rejection, that is, without correcting the likelihood ratio for its presence as in eq. (14), the false source is very readily detected. In fact, it is easier to detect than the source because its lines are sharper (due to less shielding).

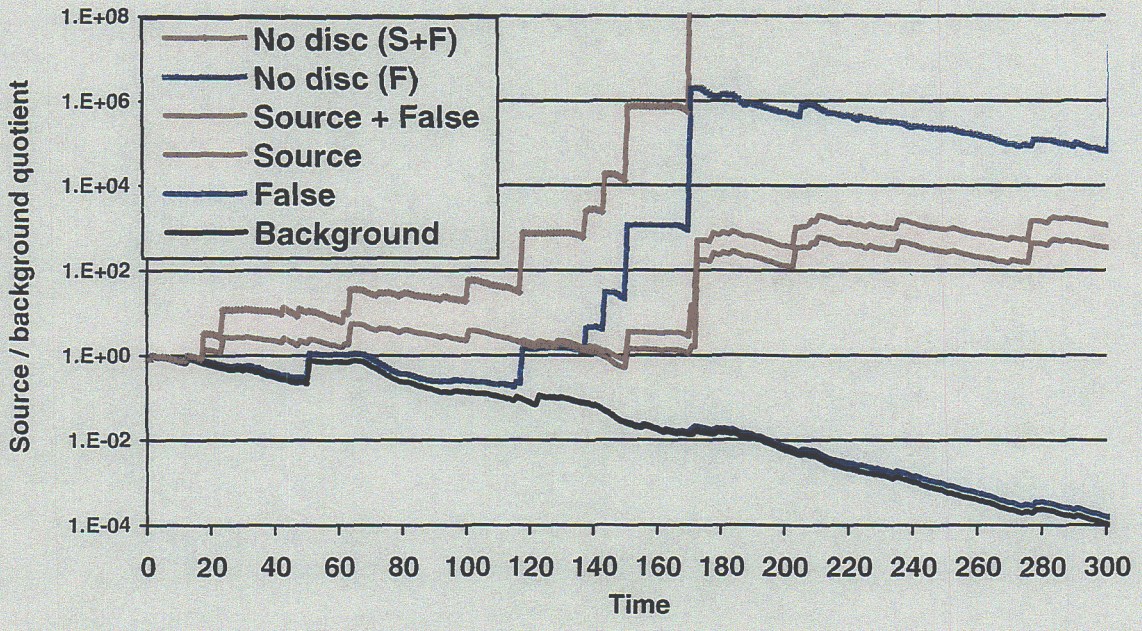

Fig. 9. Development of Q in the network with background only, false source alone, true source alone, and both true and false. Results with the discrimination capability disabled are also shown. The event at time 150, in which the false source seems to be detected slightly, involved passage of both the true and the false source by a detector simultaneously. These results are for 15-count sources.

If both the true source and the false source are present, it is necessary to reject the false source so that the true source can be tracked. In Figure 10, we disable this capability and show the resulting degradation in tracking performance.

These results show that the false-source rejection algorithm is performing exceedingly well.

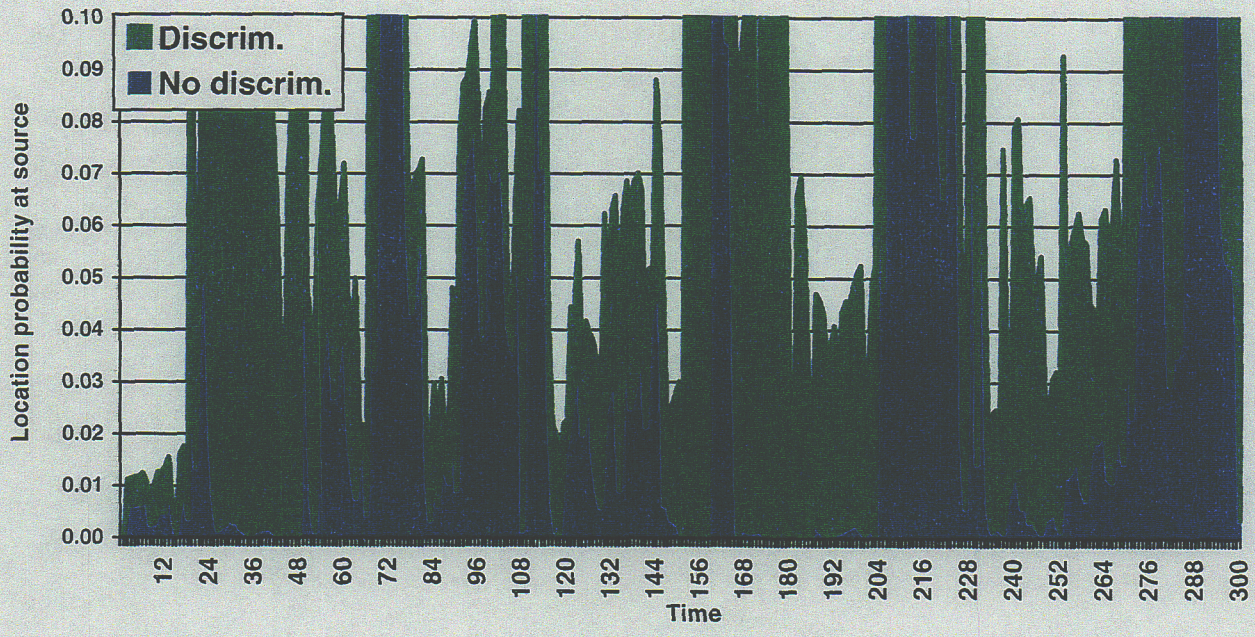

Fig.10. The ability to track the true source if the false source is present depends upon the ability to reject the false source. If this ability is compromised, the false source is tracked preferentially in our example, and the location probability estimate at the true source falls dramatically. These results are for 15-count sources.

V. Conclusions

We had three aims in this paper, to develop an algorithm for spectral classification and false-source rejection based on Bayesian techniques; to compare Bayesian techniques with current standard methods, which are largely based on a perhaps implicit reliance on Gaussian statistics; and to demonstrate their performance in a network.

It is straightforward to define the likelihood ratio for a spectrometer measurement if the spectrum to be detected and the spectrometer's background are known. This quantity is the proper basis for spectral classification and data fusion among platforms in the network. Two developments of this concept were pursued: the 'background' or null state might not be simply the instrumental background but might include expected false sources. Likelihoods for these other sources would be included in the background likelihood. Similarly, the source state to be detected might not be a pure spectrum either, but there might be several source possibilities to include in the source likelihood. We saw little performance degradation for these inclusions, but the situation of a very large number of source or background possibilities was not addressed. This remains a research issue. One technique our examples taught us to avoid was to attempt to interpret the

observations with maximum likelihood models for source or background and then use these maximum likelihoods in the likelihood ratio. Exactly why such an appealing heuristic technique fails is also a subject for further study.

We intentionally chose spectra with very low counts per channel to stress discrimination methods that rely on Gaussian statistics. Sure enough, such techniques did not perform well in our examples, compared to our Bayesian methods that used Poisson statistics. An unexpected surprise was the Mahalanobis D^2 metric that produced very nearly optimal results for our low-count spectra, yet is based on the use of a multivariate Gaussian distribution. We speculate that it forms principal axes that combine many spectral channels and so overcomes problems with low counts, but this should be investigated. The Mahalanobis technique is a bit awkward, since it involves a matrix inversion and matrix products. Fortunately, it is computationally no more difficult to use Poisson rather than Gaussian statistics when they are appropriate, and we give a generalization of the χ^2 goodness-of-fit metric that is valid for low numbers of samples.

We presented two models for Bayesian tracking networks and constructed a simple network example to illustrate the performance of our detection and discrimination techniques. We were rather pleasantly surprised to see our network performing with 15-count source strengths above a 50-count background (SNR = 1.4), but our notional sensors were relatively close together, so travel-time uncertainties were small. We gave a heuristic for estimating network performance that seemed to work well for our example. More significantly, our scheme for rejecting a known false source worked very well: network performance was indistinguishable from the no-source case when only the false source was present. Furthermore, the false-source rejection mechanism did not degrade the detection and tracking of the true source.

These results give us quite a bit of optimism regarding development of an operational capability. The next step will be to consider detection of multiple true sources against a wide variety of false sources.

ACKNOWLEDGEMENT

Special thanks to James Morgan for helping produce this paper and to Mike O'Connell for his kind suggestions and support. The tracking project has a long history: Mike Tobin, John Warhus, Arnold Warshawsky and JohnWoodworth have been instrumental in its development and in furthering my research. The general distributed Bayesian tracking architecture was developed in collaboration with Bob Bryant and Farid Dowla. Thanks to Martin Keillor for the gracious loan of his library of gamma-ray spectra – data that will be even more important in the next phase of this project.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory, under Contract No. W-7405-Eng-48.

APPENDIX A: Bayesian Decision-making

In the Bayesian approach, probabilities are assigned to individual hypotheses. Let **E** denote a statistical ensemble of possibilities (microstates), for the system that agree with our current state of knowledge of it (its macrostate). Before any measurements are made the ensemble is denoted $\mathbf{E_0}$ and its composition reflects our *a priori* knowledge of the system. For example, if we are measuring radioisotopes with a spectrometer, we may have some notion of how likely particular types might be, and this would be reflected by the relative sizes of the populations of their microstates in $\mathbf{E_0}$. As measurements are made, our knowledge of the system increases and the composition of the current ensemble changes. Let $\mathbf{E_i}$ denote the ensemble after the i^{th} measurement; its microstates agree with the *a priori* knowledge and the first i measurement results. Thus, $\mathbf{E_i} \subset \mathbf{E_{i-1}}$, since some microstates of $\mathbf{E_{i-1}}$ will not agree with the i^{th} measurement.

The probability of a hypothesis h in the ensemble $\mathbf{E_i}$ is denoted $P_i(h)$ and is the fraction of microstates in $\mathbf{E_i}$ for which h is true. Thus, $P_0(h)$ represents our *a priori* knowledge of the hypotheses; e.g. the probabilities of encountering certain isotopes in our spectroscopy example. With no subscript, P(h) refers to the ensemble for all measurements.

Since the system is being measured to validate one or more of the hypotheses, the hypotheses h must be related to measurement outcomes m. This is expressed as the conditional probability that m will be observed if h is true, $p(m \mid h)$, and this is termed the *Bayesian **likelihood** of m for* h. Note that for this likelihood to be meaningful, the measurement results must be assumed independent, provided that h is true. That is, the probability $p(m \mid h)$ cannot also depend upon previous measurement results. This in turn implies that the hypotheses h must be sufficiently specific to make them good predictors of measurement outcomes.

Since the $p(m_{i+1} \mid h)$ is by definition the fraction of the h-microstates in $\mathbf{E_i}$ that will agree with the next (i+1) measurement,

$$P_{i+1}(h) \propto P_i(h)\, p(m_{i+1} \mid h) \qquad \text{(A1)}$$

If we have a set of hypotheses {h} that are complete and disjoint, i.e. exactly one of them is true for each microstate, then we have *Bayes' Theorem:*

$$P_{i+1}(h) = \frac{P_i(h)\, p(m_{i+1} \mid h)}{\Sigma_{h'} P_i(h')\, p(m_{i+1} \mid h')} \qquad \text{(A2)}$$

However, note that eq. (1) can help us proceed for weaker assumptions on the set of hypotheses, or when normalization of probabilities may be computationally inconvenient.

As a very important example of this, note that in problems for which there is an obvious reference hypothesis, or null hypothesis n, we can define the *probability* ***quotient*** *for* h*:*

$$Q(h) \propto \quad P(h) / P(n) \tag{A3}$$

We define this quantity as a proportionality, rather than an equality, since we probably do not know its actual value initially in a detection problem.

By (1), the evolution of this quotient is

$$Q_{i+1}(h) = r(m_{i+1}, h)\, Q_i(h) \tag{A4a}$$

where r(m, h) is the *likelihood* ***ratio*** *for m given* h*:*

$$r(m, h) \equiv p(m \mid h) / p(m \mid n) \tag{A4b}$$

Since the measurement outcomes for a given hypothesis are assumed to be independent,

$$Q(h) = R(h)\, Q_0(h) \tag{A5a}$$

where R(h) is the *overall likelihood ratio given* h,

$$R(h) \equiv \Pi_i\, r(m_i, h) \tag{A5b}$$

In Appendix B, we prove that R(h) is an optimal decision variable for the hypothesis h, as opposed to the null hypothesis n, based on the entire set of measurements m. That is, the classsification scheme,

$$\text{h: for } R(h) > T; \qquad \text{n: for } R(h) < T \tag{A6}$$

where T is a variable threshold value, gives maximum probability of detection P_D with minimum probability of false alarm P_{FA}. The probability of detection is the fraction of h microstates in $\mathbf{E_0}$ that are (correctly) classified as h and the probability of false alarm is the fraction of n microstates that are (incorrectly) classified as h.

The decision scheme (A6) implies that any monotonic function of f(R(h)), is also an optimal decision variable:

$$f(R(h)) > f(T) \quad \Leftrightarrow \quad R(h) > T \tag{A7}$$

In particular Q(h) (=R(h)Q_0(h)) and Λ(h) (≡lnR(h)) are also optimal decision variables.

Classification between any two sets of hypotheses **A, B** proceeds in a similar manner. An optimal decision variable (by the proof of optimality) is

$$R(\mathbf{A},\mathbf{B}) = \Sigma_{h\in\mathbf{A}} P(h) / \Sigma_{h'\in\mathbf{B}} P(h') \qquad \text{(A8a)}$$

and a proportional, and therefore equivalently optimal, decision variable is

$$R(\mathbf{A},\mathbf{B}) = \Sigma_{h\in\mathbf{A}} f_{\mathbf{A}}(h)\, R(h) \;/\; \Sigma_{h'\in\mathbf{B}} f_{\mathbf{B}}(h)\, R(h) \qquad \text{(A8b)}$$

where $f_{\mathbf{A}}$ or $f_{\mathbf{B}}$ is the *a priori* fraction for a given hypothesis in the sets **A** or **B**:

$$f_{\mathbf{A}}(h) = P_0(h) / \Sigma_{h'\in\mathbf{A}} P_0(h') \qquad \text{(A8c)}$$

Note that for neither of the decision variables R(h) or R(**A**,**B**) do the relative probabilities of the two classes to be discriminated need to be known *a priori*. This implies, for example, that an optimal detection scheme can be devised for a rare event without having to know initially just how rare it is.

APPENDIX B: Proof of Optimality of the Likelihood Ratio as a Decision Variable

For any decision-making scheme with a well-defined null- or reference hypothesis n and a mutually-exclusive test hypothesis h, we define **n** and **h** to be the subsets of $\mathbf{E_0}$ on which n and h are true, respectively. Then the **probability of detection $\mathbf{P_D}$** is defined as the fraction of the elements of **h** that are (correctly) identified as h. The **probability of false alarm $\mathbf{P_{FA}}$** is the fraction of elements of **n** that are (mistakenly) identified as h. The assertion is that a decision-making scheme that uses the overall likelihood ratio as the decision variable will have maximum P_D at constrained P_{FA} and minimum P_{FA} at constrained P_D, compared to any other decision-making scheme that relies only on the results of measurement m_i.

If **M** denotes a set of measurement values, then a decision-making scheme based only on measurement values must be an ordered set $\mathbf{M}\equiv\{\mathbf{M}_i\}$. The hypothesis h will be declared to be true for a measurement series yielding $\mathbf{M}_i$ with $i \le I$, where I is the threshold of the scheme. Accordingly, for the threshold I

$$P_D(I) = \Sigma_{i\le I}\, p(\mathbf{M}_i \mid h) \qquad \text{(B1a)}$$

$$P_{FA}(I) = \Sigma_{i\le I}\, p(\mathbf{M}_i \mid n) \qquad \text{(B1b)}$$

since these sums represent the fractions of **h** and **n** for which h will be declared true.

The decision making scheme that we assert to be optimal has $R(\mathbf{M}_i, h) \ge R(\mathbf{M}_j, h)$ for $i < j$. If this is not true then there must be an optimal scheme $\mathbf{M}^*$ that has $R(\mathbf{M}_{i^*}, h) < R(\mathbf{M}_{i`}, h)$ for some $i^* < i`$. Suppose for the threshold $I = i^*$. Then for this scheme

$$P_D^* = \Sigma_{i<i^*} p(\mathbf{M}_i \mid h) + p(\mathbf{M}_{i^*} \mid h) \tag{B2a}$$

$$P_{FA}^* = \Sigma_{i<i^*} p(\mathbf{M}_i \mid n) + p(\mathbf{M}_{i^*} \mid n) \tag{B2b}$$

We will prove that $\mathbf{M}^*$ cannot be optimal by constructing an alternative scheme $\mathbf{M}'$ that has $P_D' > P_D^*$ for $P_{FA}' = P_{FA}^*$. There are two cases,

I. $p(\mathbf{M}_{i'} \mid n) \geq p(\mathbf{M}_{i^*} \mid n)$:

In this case, if $\mathbf{M}_{i'}$ is observed h is declared true a fraction $f = p(\mathbf{M}_{i^*} \mid n) / p(\mathbf{M}_{i'} \mid n)$ of the time, and if $\mathbf{M}_{i^*}$ is observed h is declared false. Thus for this scheme the relations $P_{FA}' = P_{FA}^*$ and $P_D' > P_D^*$ may be established:

$$P_{FA}' = \Sigma_{i<i^*} p(\mathbf{M}_i \mid n) + f\, p(\mathbf{M}_{i'} \mid n) \;=\; P_{FA}^* \tag{B3a}$$

$$P_{D'} = \Sigma_{i<i^*} p(\mathbf{M}_i \mid h) + f\, p(\mathbf{M}_{i'} \mid h)$$

$$= P_D^* + [R(\mathbf{M}_{i'}, h) - R(\mathbf{M}_{i^*}, h)]\, p(\mathbf{M}_{i^*} \mid n) \;>\; P_D^* \tag{B3b}$$

II. $p(\mathbf{M}_{i'} \mid n) < p(\mathbf{M}_{i^*} \mid n)$:

In this case, if $\mathbf{M}_{i'}$ is observed it is decided true, and if $\mathbf{M}_{i^*}$ is observed it is decided true a fraction $f = 1 - p(\mathbf{M}_{i'} \mid n) / p(\mathbf{M}_{i^*} \mid n)$ of the time, and we have, again,

$$P_{FA}' = \Sigma_{i<i^*} p(\mathbf{M}_i \mid n) + p(\mathbf{M}_{i'} \mid n) + f\, p(\mathbf{M}_{i^*} \mid n) = P_{FA}^* \tag{B4a}$$

$$P_{D'} = \Sigma_{i<i^*} p(\mathbf{M}_i \mid h) + p(\mathbf{M}_{i'} \mid h) + f\, p(\mathbf{M}_{i^*} \mid n)$$

$$= P_D^* + [R(\mathbf{M}_{i'}, h) - R(\mathbf{M}_{i^*}, h)]\, p(\mathbf{M}_{i'} \mid n) \;>\; P_D^* \tag{B4b}$$

A decision scheme that has $P_{FA}' < P_{FA}^*$ for $P_{D'} = P_D^*$ can be constructed similarly.

---

[i] Barnett, C. S.; On Some Statistical Problems Inherent In Radioactive-Source Detection; Lawrence Livermore Laboratory, UCRL-52588; October 12, 1978.

[ii] Stone, L. D., Barlow, C. A., and Corwin, T. L.; Bayesian Multiple Target Tracking; Artech House Publishers; 1999.

[iii] Cunningham, C. T.; Detection and Track of a Stochastic Target using Multiple Measurements; Lawrence Livermore National Laboratory, UCRL-ID-122786; November 1, 1995.

[iv] Mitchell, D. J.; GADRAS-95 User's Manual; Sandia National Laboratories; August 9, 1995.

[v] Gosnell, T. B., et al.; Gamma-Ray Identification of Nuclear Weapon Materials; Lawrence Livermore National Laboratory; UCRL-ID 29115; January 22, 2001.

[vi] Keillor, M. E.; Principal Component Analysis of Low Resolution Energy Spectra to Identify Gamma Sources in Moving Vehicle Traffic; Oregon State University, doctoral dissertation; September 12, 2000.

[vii] Giri, N. C.; Multivariate Statistical Inference; Academic Press, New York; 1977.

[viii] Keillor, M. E; *op. cit.*; For his dissertation, Keillor generated 181 MCNP simulations for spectra from 33 isotopes with various amounts of lead shielding, a very useful data set for classification.